

Puffer, Gabriele

## **FALKO-M: Ein Testinstrument zum domänenspezifischen Professionswissen von Musiklehrkräften**

Cvetko, Alexander J. [Hrsg.]; Rolle, Christian [Hrsg.]: *Musikpädagogik und Kulturwissenschaft*. 1. Aufl. Münster; New York : Waxmann 2017, S. 309-327. - (Musikpädagogische Forschung; 38)



Quellenangabe/ Reference:

Puffer, Gabriele: FALKO-M: Ein Testinstrument zum domänenspezifischen Professionswissen von Musiklehrkräften - In: Cvetko, Alexander J. [Hrsg.]; Rolle, Christian [Hrsg.]: *Musikpädagogik und Kulturwissenschaft*. 1. Aufl. Münster; New York : Waxmann 2017, S. 309-327 - URN: urn:nbn:de:0111-pedocs-156408 - DOI: 10.25656/01:15640

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-156408>

<https://doi.org/10.25656/01:15640>

in Kooperation mit / in cooperation with:



<http://www.ampf.info>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

*Alexander J. Cvetko, Christian Rolle (Hrsg.)*

# MUSIKPÄDAGOGIK UND KULTURWISSENSCHAFT

MUSIC EDUCATION AND  
CULTURAL STUDIES

# Musikpädagogische Forschung

## Research in Music Education

Herausgegeben vom Arbeitskreis  
Musikpädagogische Forschung e. V. (AMPF)

Band 38

Proceedings of the 38th Annual Conference of the  
German Association for Research in Music Education

Alexander J. Cvetko, Christian Rolle (Hrsg.)

# Musikpädagogik und Kulturwissenschaft

# Music Education and Cultural Studies



**Waxmann 2017**  
Münster • New York



### **Bibliografische Informationen der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

### **Musikpädagogische Forschung, Band 38 Research in Music Education, vol. 38**

ISSN 0937-3993

ISBN 978-3-8309-3661-9

© Waxmann Verlag GmbH, Münster 2017  
Steinfurter Straße 555, 48159 Münster

[www.waxmann.com](http://www.waxmann.com)  
[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Anne Breitenbach, Münster  
Satz: Heide Reinhäkel, Berlin  
Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,  
säurefrei gemäß ISO 9706



Printed in Germany

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.  
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des  
Verlages in irgendeiner Form reproduziert oder unter Verwendung  
elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

# Inhalt

*Alexander J. Cvetko & Christian Rolle*

Vorwort ..... 9

*Editors' note*

*Barbara Hornberger*

Musik – Kultur – Pädagogik

Kulturwissenschaftliche Fragen und Perspektiven ..... 19

*Music – culture – education*

*Questions and perspectives of cultural science*

*Bernd Clausen*

Regionalität und Territorialität als kulturwissenschaftliche Kategorien

historisch-musikpädagogischer Forschung am Beispiel des Tonwortstreits:

Der Fall Raimund Heuler ..... 37

*Regionality and territoriality as categories of cultural studies*

*in historical research in music education: The latonisation dispute and*

*Raimund Heuler*

*Heinz Geuen & Christine Stöger*

„Spielarten“ – Musizieren im allgemeinbildenden Musikunterricht

aus Perspektive der Cultural Studies: Ein Gedankenexperiment ..... 57

*“Spielarten” – Making music in general music education from a cultural studies*

*perspective: A thought experiment*

*Marc Godau & Matthias Krebs*

Weiterbildung als Community of Practice?

Zur forschungsbasierten Entwicklung der Weiterbildung

„Zertifikatskurs tAPP – Musik mit Apps in der Kulturellen Bildung“ ..... 73

*Further education as community of practice? Insights into the*

*research-based development of the “certificate course tAPP –*

*music with apps in cultural education”*

*Valerie Krupp-Schleußner*

„Kopfhörer rein, Welt aus.“ Individuelle Perspektiven auf musikalisch-kulturelle Teilhabe vor dem Hintergrund des *capability approach* ..... 87

*“Put on your headphones; tune out the world.” Individual perspectives on cultural participation in light of the capability approach*

*Andreas Kruse*

Wider eine Ästhetik (im Dienst) der Gewalt

Kulturwissenschaftliche Impulse zu einem musikpädagogischen Problem ..... 105

*Opposing aesthetics that serve violent ends: The impulses of cultural studies to move toward a music-educational approach to challenges*

*Stefan Orgass*

Ein System als Teil einer Theorie

Überlegungen zum Abgrenzungskriterium einer reflexionslogischen

Wissenschaftstheorie der Musikpädagogik ..... 119

*A system as part of a theory: Reflections on the problem of demarcation as a problem of scientific music pedagogy*

*Michael Rappe & Christine Stöger*

„I wanna do b-boy moves, but I wanna be known as a b-girl“

Breaking lernen als Prozess der Identitätskonstruktion ..... 137

*“I wanna do b-boy moves, but I wanna be known as a b-girl”*

*Learning to break as an identity construction process*

*Christian Rolle, Elin Angelo & Eva Georgii-Hemming*

Mapping the methodological field of discourse analysis

in music education research

A review study, part I ..... 153

*Constanze Rora*

Musik als Praxis aus dem Blickwinkel einer Phänomenologie

der Partizipation ..... 165

*Music as human practice in the perspective of phenomenological participation*

*Peter W. Schatt*

„... das Wörtlein: und“: Überlegungen zur Erkundung und Gestaltung  
des Verhältnisses zwischen Musikpädagogik und Kulturwissenschaft ..... 181

*„...das Wörtlein: und“: Reflections on the exploration and development  
of the relationship between music education and cultural studies*

## **Freie Beiträge**

*Thomas Busch*

Die Entwicklung der Selbstregulation instrumentalen Übens  
in der späten Kindheit ..... 199

*The development of self-regulation in instrumental practice during late childhood*

*Daniel Fiedler & Daniel Müllensiefen*

Musikalisches Selbstkonzept und Musikalische Erfahrungheit beeinflussen  
die Entwicklung des Interesses am Schulfach Musik  
Eine empirische Längsschnittuntersuchung von Schülerinnen und Schülern  
an Haupt-, Gemeinschafts- und Realschulen sowie Gymnasien in Baden-  
Württemberg ..... 217

*Musical self-concept and musical sophistication influence the development of  
interest in music as a school subject. An empirical long-term study of German  
students at secondary schools*

*Marc Godau*

Apps in der musikpädagogischen Praxis  
Eine explorative Studie zur kommunikativen Konstruktion von mobilen  
Technologien im schulischen Nachmittagsbereich ..... 237

*Music education practices with apps: Exploring the social construction of mobile  
technology in extracurricular classes*

*Michael Göllner*

Perspektiven von Lehrenden sowie Schülerinnen und Schülern auf  
das Vexierbild Bläserklassenunterricht  
Eine qualitative Interviewstudie ..... 251

*Perspectives of teachers and pupils on the concert band as a picture puzzle  
A qualitative interview study*



<i>Christian Harnischmacher, Kilian Blum &amp; Viola Cäcilia Hofbauer</i>	
Kompetenzorientierung und Motivation von Lehrern im Musikunterricht . . . . .	269
<i>Competence orientation and motivation of teachers in music lessons</i>	
 <i>Daniel Prantl &amp; Christopher Wallbaum</i>	
Der Analytical Short Film in der Lehrerbildung	
Darstellung einer Seminarmethode und Kurzbericht einer wissenschaftlichen	
Begleitforschung an der Hochschule für Musik und Theater Leipzig . . . . .	289
<i>The analytical short film in teacher training – Presentation of</i>	
<i>a seminar method and short report on the accompanying research</i>	
<i>at the University of Music and Drama, Leipzig</i>	
 <i>Gabriele Puffer</i>	
FALKO-M: Ein Testinstrument zum domänenspezifischen Professionswissen	
von Musiklehrkräften . . . . .	309
<i>FALKO-M: A tool for assessing music teachers' professional knowledge</i>	
 <i>Jürgen Vogt</i>	
Versuch über Kritische Musikpädagogik . . . . .	329
<i>An essay on critical music education</i>	

Gabriele Puffer

## **FALKO-M: Ein Testinstrument zum domänenspezifischen Professionswissen von Musiklehrkräften**

*FALKO-M: A tool for assessing music teachers' professional knowledge*

*This paper reports on the results of a test development process that was conducted as part of the interdisciplinary FALKO project. Based on Lee Shulman's theoretical taxonomy of teacher knowledge and the preliminary work of the COACTIV research program, six domain specific tests for assessing secondary school teachers' content knowledge (CK) and pedagogical content knowledge (PCK) have been constructed and validated for the subjects German, English, Religious Education, Latin, Music Education and Physics. Results of the validation study provide evidence that Shulman's categories CK and PCK are applicable to the professional knowledge of general music teachers. Furthermore, reliability and validity of the FALKO-M test can be considered as sufficient to permit its use in future research projects.*

### **Einleitung**

Wann und wo erwerben Lehrkräfte ihr professionelles Wissen? Gibt es Ausbildungsphasen mit besonders großen Wissenszuwächsen? Folgt der Wissensaufbau einer Lehrkraft in allen ihren Unterrichtsfächern einem ähnlichen Verlauf, oder gibt es unterschiedliche fachspezifische Entwicklungsmuster? Um solchen Fragen nachzugehen, benötigt man ein Forschungswerkzeug, das das Untersuchen fachspezifischer Zusammenhänge und Entwicklungsverläufe ebenso erlaubt wie fachübergreifende Vergleiche. Aus diesem Grund entwickelte die interdisziplinäre Forschungsgruppe FALKO an den Universitäten Regensburg und Augsburg sechs Tests zu Fachwissen und fachdidaktischem Wissen in den Unterrichtsfächern Deutsch, Englisch, Evangelische Religionslehre, Latein, Musik und Physik (Krauss et al., 2017). Zielgruppe sind jeweils Lehrkräfte der Sekundarstufe I und Studierende aus entsprechenden Studiengängen. Jeder Test ist fachspezifisch ausgelegt und kann separat eingesetzt werden; ein einheitliches theoretisches und formales Rahmenkonzept ermöglicht aber auch das Prüfen fächerübergreifender Zusammenhangs- und Unterschiedshypothesen (erste transdisziplinäre Ergebnisse in Lindl & Krauss, 2017).



Relevanz erhält die Untersuchung fachspezifischen Professionswissens vor dem Hintergrund der Vorstellung von Lehrerwissen als einem wirkmächtigen, prinzipiell erlern- und vermittelbaren Faktor professioneller Kompetenz von Lehrkräften (neben Aspekten wie berufsbezogenen Überzeugungen, motivationalen Orientierungen, Fähigkeiten zur Selbstregulation; vgl. Baumert & Kunter, 2011a; Voss, Kunina-Habenicht, Hoehne & Kunter, 2015, S. 189f.; Mulder & Gruber, 2011). Anknüpfend an Lee Shulmans Taxonomie von Lehrerwissen aus den 1980er-Jahren (Shulman, 1986, 1987) sind in der deutschsprachigen empirischen Bildungsforschung die Facetten Fachwissen, fachdidaktisches Wissen und allgemein-pädagogisches Wissen von Lehrkräften mittlerweile als konzeptuell und empirisch trennbare „Kernkategorien des Lehrerwissens“ etabliert (Krauss et al., 2017, S. 16 und 26; Voss et al., 2015). Einen weiteren Bezugspunkt des Projekts FALKO bilden Vorarbeiten des Forschungsprogramms COACTIV (Kunter et al., 2011): Hier konnte zum einen ein erheblicher Einfluss des fachdidaktischen Wissens von Mathematiklehrkräften auf Unterrichtsqualität und Lernzuwachs von Schülerinnen und Schülern nachgewiesen werden (Baumert & Kunter, 2011b). Zum anderen verwiesen deutliche quantitative Unterschiede zwischen berufserfahrenen Lehrkräften und Studierenden sowie zwischen Teilnehmenden aus gymnasialen und nicht gymnasialen Studiengängen auf die Ausbildungsabhängigkeit professioneller Wissensbestände (Baumert & Kunter, 2011b). Dies wirft die Frage auf, ob sich ähnliche Zusammenhänge auch für andere Fächer feststellen lassen (Krauss et al., 2017, S. 12). Zwei zentrale Hypothesen der Testentwicklung im Projekt FALKO beziehen sich deshalb auf die Ausbildungsabhängigkeit von Fachwissen und fachdidaktischem Wissen: Professionelles Wissen sollte im Laufe der Lehramtsausbildung und auch während der späteren Berufsausübung zunehmen (vgl. Krauss et al., 2008, S. 244; Mulder & Gruber, 2011). Ein valider, handlungsnah konzipierter Professionswissenstest müsste entsprechende Unterschiede zwischen berufserfahrenen Lehrkräften und Studierenden abbilden. Geht man zudem davon aus, dass sich eine fachlich intensivere Ausbildung in größeren Beständen professionellen Wissens niederschlägt, sollten sich zudem für alle beteiligten Unterrichtsfächer quantitative Unterschiede zwischen Lehrkräften und Studierenden gymnasialer und nicht gymnasialer Schulformen messen lassen – zumindest für den Bereich des Fachwissens (Krauss et al., 2017, S. 12).

## Konzeption des Tests

Der hier vorzustellende Professionswissenstest soll für Musiklehrkräfte aller Sekundarstufenlehrämter und Ausbildungsniveaus gleichermaßen bearbeitbar sein. Deshalb wurde er mit Blick auf Anforderungen des Musikunterrichts in den Klassenstufen 5 und 6 konzipiert. In diesen beiden Jahrgängen ist über alle Schularten hinweg die Wahrscheinlichkeit am größten, dass Musikunterricht auch



tatsächlich erteilt wird (vgl. die Überlegungen in Niessen, Lehmann-Wermser, Knigge & Lehmann, 2008, S. 18; Jank, 2013, S. 86f.). Die inhaltliche Fokussierung erhöht also die Anzahl potenzieller Testteilnehmerinnen und -teilnehmer, die auch tatsächlich über einschlägige Unterrichtserfahrung verfügen.

Zu entwickeln war ein Papier-und-Bleistift-Test mit möglichst unterrichtsnaher inhaltlicher Ausrichtung und einer Mischung geschlossener und offener Itemformate, der von Lehrkräften aller infrage kommenden Schularten in einem zumutbaren Zeitrahmen von maximal 90 Minuten bearbeitbar sein sollte. Um alle konzeptualisierten Facetten des fachdidaktischen Wissens angemessen abbilden zu können, war fachübergreifend eine Arbeitszeit von etwa 60 Minuten für den Subtest „Fachdidaktisches Wissen“ angestrebt; für den Subtest „Fachwissen“ verblieben damit ca. 30 Minuten. Professionelles Wissen von Musiklehrkräften innerhalb dieses Rahmens in akzeptabler Breite abzubilden erwies sich in vielerlei Hinsicht als Herausforderung: Die fachtypische Pluralität didaktischer Modelle und Konzepte sowie heterogene curriculare Vorgaben aus 16 Bundesländern machen schulischen Musikunterricht im Sinne der Expertiseforschung zu einer gering strukturierten Domäne. Hinzu kommt die große Bandbreite von Ausbildungskonzepten und -niveaus in der Musiklehrerbildung. Die Entwicklung empirisch fundierter musikdidaktischer Theorien zum Thema Lehrerkompetenz steckt noch in den Anfängen. Im Zuge der Testentwicklung mussten daher zahlreiche vorläufige Setzungen vorgenommen werden (vgl. Puffer & Hofmann, 2016). Dies betrifft insbesondere die Frage, wie und wann Anforderungen der schulischen Unterrichtspraxis adäquat bewältigt sind.

Bei der Entwicklung der Testitems und der Bewertungskriterien im Kodiermanual kam deshalb den Rückmeldungen von Expertinnen und Experten aus Schulen und Hochschulen eine bedeutende Rolle zu. Die Items des Professionswissenstests wurden zunächst auf der Basis von Literaturrecherchen zusammengestellt (schulische Curricula, Standards für die Ausbildung von Musiklehrkräften, musikdidaktische Literatur, Unterrichtswerke, Handreichungen für Lehrkräfte, Fachzeitschriften) und anschließend mit Fachkolleginnen und -kollegen, Studierenden und erfahrenen Lehrkräften diskutiert. Besonderen Stellenwert hatte dabei das Votum der Lehrkräfte: Die Testaufgaben sollten inhaltlich nicht nur musikdidaktische Lehrmeinung repräsentieren, sondern möglichst nah an Erfordernissen der Unterrichtspraxis angesiedelt sein.

In die Endfassung des Tests gingen nach mehreren Pilotierungsrunden 11 Items zum fachdidaktischen Wissen und 8 Fachwissensitems ein. Beibehalten wurden Items, die im Rahmen eines Expertenratings als relevant für die Ausbildung von Musiklehrkräften eingestuft wurden, was sich als Indiz für ökologische Validität interpretieren lässt,<sup>1</sup> von allen Pilotierungsteilnehmenden im vorgese-

1 Im Rahmen der Prä-Pilotierung beurteilten Musikdidaktikerinnen und -didaktiker, Seminarlehrkräfte, Fachberater sowie Fachbetreuerinnen und -betreuer aus dem Hauptschul-, Realschul- und Gymnasialbereich ( $N = 29$ ) jedes Item bezüglich der Verständlich-



nenen Zeitrahmen bearbeitet werden konnten, sich mit guter Übereinstimmung jeweils zweier Rater kodieren ließen (angestrebt war ein Wert von Spearmans  $\rho \geq .90$ ) und befriedigende Trennschärfen bezüglich der jeweils zugewiesenen Wissensfacette aufwiesen ( $r_{it} \geq .20$ ).

## Fachwissen

Die Items des Subtests „Fachwissen“ zielen auf Bestände vertieften Wissens über curriculare Inhalte des Musikunterrichts in den Jahrgangsstufen 5 und 6. Damit der FALKO-M-Test über einen längeren Zeitraum hinweg einsetzbar bleibt, wurde eine Auswahl vergleichsweise zeitloser Themen vorgenommen: Repertoirekenntnisse und Hintergrundwissen über Werke „klassischer“ Musik und Rock-/Popsongs, die in Lehrwerken und Handreichungen für den Musikunterricht oder in Fachzeitschriften besonders häufig und über einen längeren Zeitraum hinweg anzutreffen sind; Grundlagenwissen über Stimmbildung, Stimmpflege und Liedbegleitung, musikalische Fachbegriffe und Notation von Musik sowie Weltmusik-Instrumentarium; Umgang mit technischem Equipment (Audio-Mischpult). Da hier in erster Linie Faktenwissen verlangt wird, finden fast ausschließlich geschlossene und teilstrukturierte Aufgabenformate wie in Abbildung 1 Verwendung: Single-Choice-, Multiple-Choice- sowie Ergänzungsaufgaben, für die es jeweils genau eine korrekte Lösung gibt (im Falle der Aufgabe in Abbildung 1: c – as’).

## Fachdidaktisches Wissen

Analog zum Vorgehen des COACTIV-Forschungsteams und unter Berücksichtigung fachspezifischer Anforderungen ist in FALKO-M die fachdidaktische Wissensgrundlage der Lehrkräfte in drei Facetten konzeptualisiert:

- Wissen über musikbezogene Instruktionsstrategien,
- Wissen über musikbezogene Kognitionen und Lernwege von Schülerinnen und Schülern,
- Wissen über das Potenzial von Materialien für musikbezogene Lehr-/Lernprozesse.

---

keit der Aufgabenstellung sowie der Relevanz des jeweiligen Inhalts für die Ausbildung von Musiklehrkräften. Die Bewertung wurde jeweils mithilfe einer vierstufigen Likert-Skala vorgenommen (von 1 = „unverständlich“ bis 4 = „sehr gut verständlich“ beziehungsweise 1 = „nicht relevant“ bis 4 = „sehr relevant“). Im Auswahlverfahren blieben Items mit einer mittleren Bewertung von  $M > 2.5$ .

Das folgende Notenbeispiel stammt aus „Le Carnaval des Animaux“ („Der Karneval der Tiere“) von Camille Saint-Saëns:

Clarinetten in Si<sup>b</sup>  
(dans la coulisse)

Andante

*f*

Piano I  
una corda *pp*

Piano II  
una corda *pp*

### Klarinettenstimme

Die in der Klarinettenstimme dieses Beispiels notierten Töne erklingen als (Zutreffendes bitte ankreuzen)...

- ☐ d<sup>''</sup> – b<sup>'</sup>
- ☐ e<sup>''</sup> – c<sup>''</sup>
- ☐ c<sup>''</sup> – as<sup>'</sup>
- ☐ b<sup>'</sup> – ges<sup>'</sup>

Abbildung 1: Beispiel „Karneval der Tiere“; Aufgabenstamm und Beispielitem

Als weiterer heuristischer Hintergrund zur inhaltlichen Strukturierung des Subtests dienen die von Dankmar Venus formulierten fünf *vorrangigen Verhaltensweisen* gegenüber Musik (Venus, 1969), im musikpädagogischen Diskurs häufig als *Umgangsweisen* bezeichnet (vgl. Kaiser & Nolte, 1989, S. 31f.). Diese Entscheidung gründet sich darauf, dass Venus' Kategorisierung bis heute schulische Curricula und theoretische Modelle grundiert, beispielsweise in Form von Lernbereichen des Musikunterrichts oder als Raster, auf dessen Grundlage Dimensionen musikalischer Kompetenz modelliert werden (vgl. hierzu Kaiser & Nolte, 1989, S. 30–34; Gruhn, 2003, S. 328f.; Kraemer, 2007, S. 179–185; Niessen et al., 2008, S. 10; Jank, 2013, S. 40f. und S. 125). Im FALKO-M-Test sind alle Umgangsweisen repräsentiert, mit einem Schwerpunkt auf Reproduktion (vgl. Tabelle 1). Damit wird dem Anliegen Rechnung getragen, (gemeinsame) musikalische Praxis im Test als besonders bedeutsames Moment der Unterrichtsgestaltung zu berücksichtigen (Puffer & Hofmann, 2016). Im Rahmen einer handlungsnah ausgerichteten Testkonzeption erwies es sich allerdings als nicht sinnvoll, jedes Testitem genau einer Umgangsweise zuzuordnen: Hier wurden psychometrische Kriterien zugunsten inhaltlicher Validität hintangestellt. Umgangsweisen mit Musik lassen sich zwar auf analytischer Ebene voneinander trennen, nicht aber im Vollzug von Musiklehren und -lernen (Kraemer, 2007, S. 182). Zudem benennen musikdidaktische Schriften und schulische Curricula die sinnvolle Vernetzung verschiedener Umgangsweisen als Gütekriterium eines zeitgemäßen Musikunterrichts (z. B. Jank, 2013, S. 116–118; Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2014, S. 113–115).



Tabelle 1: Übersicht über Verteilung der Items (Wissensfacette x heuristischer Hintergrund) und Antwortformate. Die kursiv gedruckten Zahlen bezeichnen die von 1 bis 11 durchnummerierten Items des Subtests „Fachdidaktisches Wissen“.

Facetten	Itemanzahl (offen/ geschlossen)	1 Repro- duktion	2 Produktion	3 Rezeption	4 Transfor- mation	5 Reflexi- on
Fachwissen	8 (2/6)					
Fachdidaktisches Wissen	11 (9/2)					
Instruktion	5 (3/2)	2, 3, 6	5, 8	5	5	5
Lernwege	3 (3/0)	1, 7		7		4, 7
Potential	3 (3/0)	9, 10, 11	10	9, 10	11	9, 10, 11
$\Sigma$	19 (11/8)					

An Stelle des von Venus ursprünglich verwendeten Begriffs *Transposition* wird die weniger missverständliche Variante „Transformation“ verwendet (vgl. Kaiser & Nolte, 1989, S. 31).

Im Unterschied zum Fachwissenstest ist der Subtest „Fachdidaktisches Wissen“ überwiegend mit offenen Itemformaten gestaltet, denen eine einheitliche Konstruktionsvorschrift zugrunde liegt (vgl. Abbildung 2): Einer knappen Überschrift folgt eine kurze Sachinformation oder ein Stimulus, der visueller oder akustischer Natur sein kann (z. B. ein Ausschnitt aus einem Schulbuch für den Musikunterricht oder eine Audioaufnahme einer singenden Kindergruppe). Dem Material voran- oder nachgestellt ist die Schilderung eines Unterrichtsszenarios, gefolgt von der eigentlichen Frage oder Aufgabenstellung. Dieses Konzept soll der Kontextabhängigkeit von Kompetenzen und der dazu gehörenden Wissensbestände gerecht werden (vgl. Krauss et al., 2011, S. 138; Mulder & Gruber, 2011, S. 233; Mayer & Wellnitz, 2014, S. 27). Die Textanteile der Aufgabenstämme sind möglichst kurz gehalten, um nicht in unnötigem Umfang Lesefähigkeiten mitzutesten.

<b>„Guten Morgen“</b> In Hörbeispiel 1 singt der Kinderchor eines Kollegen.	
Welche Aspekte dieser Darbietung beurteilen Sie unter stimmbildnerischer Perspektive als besonders problematisch?	Wie würden Sie diesen Problemen in Ihrem Unterricht abhelfen? Skizzieren Sie bitte stichpunktartig die Methoden, die Sie in solchen Fällen bevorzugt verwenden!

Abbildung 2: Beispielitem „Guten Morgen“; Aufgabenstamm mit zwei Items zu den Facetten „Wissen über musikbezogene Kognitionen und Lernwege von Schülerinnen und Schülern“ und „Wissen über musikbezogene Instruktionsstrategien“

<b>„Guten Morgen“</b>			
In Hörbeispiel 1 singt der Kinderchor eines Kollegen. Welche Aspekte dieser Darbietung beurteilen Sie unter stimmbildnerischer Perspektive als besonders problematisch?			
<b>Inhaltliche Leistung</b>	<b>Beispiel</b>	<b>Antwort-code</b>	<b>Score</b>
<i>Keine Antwort</i>	Gar nichts, Strich oder „?“	999	0
<i>Falsche/unverständliche/unvollständige Lösung bzw. Aufgabe missverstanden</i>	Frage missverstanden: Teilnehmer(in) zählt z. B. Fehler auf, die nichts mit Stimmgebung zu tun haben (z. B. „große Sprünge in der Melodie“); falsche Lösung: z. B. „Instrumentalbegleitung kaum hörbar“; Mittel zur Abhilfe werden genannt (die in die Lösung zur zweiten Teilaufgabe gehören), z. B. „Einstieg von oben“, „auf schlankere Stimmgebung hin arbeiten“	0	0
<b>„Power-Aufgabe“ mit nach oben offener Höchstpunktzahl!</b>			
<i>Richtige Lösungen: Kategorien</i>	<i>Ankerbeispiele</i>		Score = Anzahl der unter Antwortcode 1-7 richtig benannten Einzelaspekte; pro Code können auch mehrere Punkte vergeben werden!
Stimmphysiologische Aspekte  <i>(Fachbegriffe müssen nicht in dieser Form genannt werden, Umschreibung genügt)</i>	Hypertoner bzw. überspannter Zustand; Singen mit (zu) viel Druck oder Anstrengung; Fixierung im Brustregister; forcierte Stimmgebung in der Mittellage, sehr lautstarkes Singen, stellenweise fast geschrien;  grelle, plärrige, flache, enge, gepresste Stimmgebung, vor allem bei den höheren Tönen;  kein Register- bzw. Lagenausgleich; Glottisschläge bei Vokalansätzen	1	
Intonationsmängel (aufgrund stimmtechnischer Defizite)	„Intonation“ (allgemein), oder konkreter, z. B.: „Sextsprung abwärts nicht getroffen (gis-h)“	2	
Artikulation	übertrieben akkurate Nachsilben; „unnatürliche“ Aussprache, überhelle Vokale; z.T. schlecht verständlich	3	
Modulation, Ausdruck	Wenig modulierter Einsatz der Stimmen; monotoner Stimmklang; keine Beziehung zwischen Ausdruck und Text	4	
Tonart	Tiefe Töne klingen dünn, da zu tiefe Lage; für Kinderstimmen recht tiefe Tonart	5	

Abbildung 3: Auszug aus dem Kodiermanual zum Beispiel „Guten Morgen“, Item 1 (Facette „Wissen über musikbezogene Kognitionen und Lernwege von Schülerinnen und Schülern“)



Aus der Formulierung der Aufgabenstellungen wird deutlich, dass es im Subtest „Fachdidaktisches Wissen“ um eine andere Art von Lehrerwissen geht als im Fachwissenstest. Hier ist eine Form strategischen Wissens gefragt, das neben Faktenwissen auch Überzeugungen der Lehrkraft einschließt (Shulman, 1986, S. 9f.). Angesichts dieser subjektiven Komponente lassen sich die meisten Testantworten nicht per se als richtig oder falsch bewerten; grundlegendes Beurteilungskriterium in allen FALKO-Tests ist didaktische Angemessenheit. Das entsprechende Kodiermanual wurde in einem mehrstufigen Prozess unter Einbeziehung von Expertenurteilen sowie auf Basis deduktiver wie induktiver Vorgehensweisen entwickelt und enthält für jedes Item ein spezifisches Kategorienschema. Die Ermittlung des jeweiligen Scores erfolgt über die Summe der als didaktisch angemessen bewerteten, substantiell unterschiedlichen Teilantworten (Puffer & Hofmann, 2017; vgl. auch Krauss et al., 2011, S. 144–146).

## Validierungsstudie


### Stichprobe und Durchführung

Die Datenerhebung nahm aufgrund restriktiver Rahmenbedingungen für Erhebungen an Schulen in Bayern den Zeitraum zwischen Juli 2013 und Oktober 2016 in Anspruch. Kriterien für die Auswahl der Studienteilnehmer waren in erster Linie deren Verfügbarkeit und Bereitschaft zur Mitwirkung. Es handelt sich also um eine Gelegenheitsstichprobe, deren Zusammensetzung bei der Dateninterpretation stets sorgfältig berücksichtigt werden muss. Die 17 teilnehmenden Referendare und Referendarinnen wurden zur Gruppe der Lehrkräfte gezählt, was sich gleichermaßen mit theoretischen Vorüberlegungen wie mit den erzielten Testleistungen begründen lässt (Puffer & Hofmann, 2017, S. 266).

Die Erhebungen wurden unter Aufsicht eines Testleiters an verschiedenen bayerischen Schulen und Hochschulen durchgeführt. Die Teilnahme an der Studie war stets freiwillig und wurde – mit Ausnahme einiger Nacherhebungen an Studierenden – nicht vergütet. Ergänzend wurden 45 Testhefte postalisch verschickt und die Hörbeispiele online bereitgestellt. Eine Überprüfung der auf dieser Basis erzielten Testleistungen ergab keine signifikanten Unterschiede zu den Ergebnissen der Erhebungen unter Anwesenheit eines Testleiters.<sup>2</sup> Daher wurden die so gewonnenen Daten in die Hauptstudie übernommen.

2 Vergleich mittels t-Tests; verglichen wurden die Summenscores für Fachwissen und fachdidaktisches Wissen, getrennt nach Schulform und beruflichem Status (Studierende GY versus Studierende GY, Studierende NGY versus Studierende NGY usw.)

Tabelle 2: Zusammensetzung der Stichprobe der FALKO-M-Validierungsstudie

	Studierende	Lehrkräfte	$\Sigma$
GY	30	58	88
NGY	21 (9 HS/MS, 12 RS)	20 (5 HS/MS, 15 RS)	41
$\Sigma$	51	78	129
	47,1%	44,9%	
Alter (in Jahren) <i>M (SD)</i>	23.67 (3.56)	40.19 (11.29)	
Studien- (Semester) bzw. Berufserfahrung (Jahre) <i>M (SD)</i>	5.08 (3.16)	11.44 (10.34)	

*M*: arithm. Mittel; *SD*: Standardabweichung; GY: Gymnasiales Lehramt, NGY: nicht gymnasiales Lehramt; HS/MS: Lehramt Haupt- beziehungsweise Mittelschule; RS: Lehramt Realschule

## Psychometrische Gütekriterien

Die interne Konsistenz der beiden Skalen wurde mithilfe von Cronbachs  $\alpha$  bestimmt und betrug für den Subtest „Fachdidaktisches Wissen“ (FDW)  $\alpha = .80$ , für den Fachwissenstest (FW)  $\alpha = .69$ . Diese Werte sind – im Rahmen einer Testneukonstruktion und unter Berücksichtigung der geringen Itemzahl – als akzeptabel bis gut einzustufen (Bühner, 2011, S. 80f.). Ebenso verhält es sich mit den Werten für die Trennschärfen.<sup>3</sup> Der vergleichsweise niedrige Wert  $r_{it} = .20$  im Subtest „Fachdidaktisches Wissen“ ist auf einen Ausreißer nach unten zurückzuführen: Das Item „Brummer“ wurde trotz vergleichsweise schlechter Kennwerte im Test belassen, da es einen nach übereinstimmender Einschätzung von Entwicklungsteam und Testteilnehmenden inhaltlich sehr relevanten Aspekt von Musikunterricht thematisiert (Augenscheinvalidität des Items im Urteil der Teilnehmerinnen und Teilnehmer der Validierungsstudie ( $N = 129$ ):  $M = 3.68$ ,  $SD = .50$ ). Die übrigen offenen Items des FDW-Tests weisen Trennschärfen von  $r_{it} \geq .41$  auf.<sup>4</sup>

Einen Anhaltspunkt zur inhaltlichen Validierung des FALKO-M-Tests bieten die erhobenen Daten zur Augenscheinvalidität. Wie die Expertinnen und Experten in der Präpilottierungsphase wurden auch die Teilnehmerinnen und Teilneh-

3 Die Trennschärfe  $r_{it}$  eines Items gibt an, wie repräsentativ es für die Gesamtskala ist (Bühner, 2011, S. 171).

4 Detaillierte deskriptive und psychometrische Kennwerte auf Itemebene in Puffer & Hofmann, 2017, S. 288f.



mer der Validierungsstudie um eine Bewertung der Ausbildungsrelevanz jedes Items gebeten. Sie stuften die Aufgabeninhalte beider Tests durchschnittlich als „ziemlich relevant“ für die Ausbildung von Musiklehrkräften ein. Der Test zum fachdidaktischen Wissen schnitt dabei etwas besser ab ( $M = 3.33$ ,  $SD = 0.32$ ) als der Fachwissenstest ( $M = 3.02$ ,  $SD = 0.19$ ).

Alle Testhefte der Pilotierung sowie 50 der 129 Testhefte der Hauptstudie wurden jeweils doppelt kodiert, woran insgesamt sechs Raterinnen und Rater in wechselnden Zweiertteams beteiligt waren. Für die Validierungsstudie konnte dabei eine äußerst zufriedenstellende Auswertungsobjektivität erzielt werden: Die mittlere Interraterreliabilität für die offenen Items (Spearman's  $\rho$ ) lag über  $\rho = .90$ , der geringste Wert betrug  $\rho = .88$ . Die in FALKO-M verwendete Kombination aus Testheft, Kodiermanual und Raterschulung erwies sich somit als geeignete Grundlage für eine objektive Testauswertung.

Tabelle 3: Reliabilität der Gesamtskalen (Cronbachs  $\alpha$ ), Trennschärfen ( $r_t$ ), Augenscheininvalidität (Ausbildungsrelevanz), Interraterreliabilität (Spearman's  $\rho$ ) und Itemschwierigkeiten (nach Dahl, 1971; vgl. Bortz & Döring, 2006, S. 219)

Skala (Item- zahl)	Interne Konsistenz $N = 129$	Trennschärfe	Augenscheininvalidität (Ausbildungs- relevanz) $N = 129$	Interraterreli- abilität $n = 50$ (doppelt kodierte Test- hefte)	Item- schwie- rigkei- ten $N = 129$ $M (SD)$
	Cronbachs $\alpha$	$r_t$	$M (SD)$	Spearman's $\rho$	
Fach- wissen (8)	.69	range [.30-.42]	3.02 (.19) range [2.74-3.38]	range [.89-.96] (2 offene Items)	.54 (.09) range [.44-. 71]
Fach- didakt. Wissen (11)	.80	range [.20-.58]	3.33 (.32) range [2.72-3.71]	range [.88-.98] (9 offene Items)	.42 (.14) range [.30- .75]

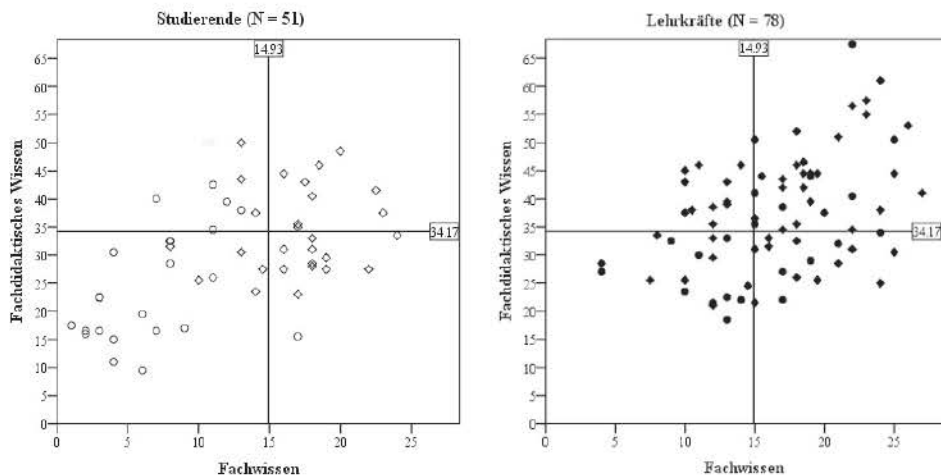
$M$ : arithm. Mittel;  $SD$ : Standardabweichung; Skalierung Augenscheininvalidität (Ausbildungsrelevanz): 1 Nicht relevant; 2 Kaum relevant; 3 Ziemlich relevant; 4 Sehr relevant.

## Deskriptive Ergebnisse

Insgesamt erscheinen die Schwierigkeiten der Items als der Zielgruppe angemessen (vgl. Tabelle 3), die Summenscores für fachdidaktisches Wissen und Fachwissen sind annähernd normalverteilt. Die Bandbreite der erzielten Testleistungen variiert erheblich, was sich als Hinweis darauf interpretieren lässt, dass beide Subtests gut zwischen schwächeren und leistungsfähigeren Studienteilnehmerinnen und -teilnehmern differenzieren: Das schlechteste Ergebnis im

Fachwissenstest lag bei einem Punkt, das beste bei 27 Punkten ( $M = 14.93$ ,  $SD = 6.02$ ). Im Subtest „Fachdidaktisches Wissen“ bewegten sich die Gesamtscores zwischen 9,50 und 67,50 Punkten ( $M = 34.17$ ,  $SD = 10.79$ ; detaillierte Aufstellung aller Kennzahlen in Puffer & Hofmann, 2017).

Grundlegend für die FALKO-Studie sind Befunde aus der Expertiseforschung. Ihnen zufolge müssten im Berufsleben stehende Lehrkräfte über einen größeren, durch berufliche Erfahrung weiterentwickelten Wissensschatz zu den Kerntätigkeiten des Lehrens verfügen als Studierende (Kunter et al., 2011, S. 63f.; Gruber & Stöger, 2011, S. 257), was sich in den Testergebnissen entsprechend niederschlagen sollte. Tatsächlich ergibt sich für die Skala „Fachdidaktisches Wissen“ ein mittlerer bis großer Effekt zugunsten der Lehrkräfte ( $t(127) = 3.89$ ,  $p < .001$ ,  $d = .70$ ), was als Indiz für Konstruktvalidität gedeutet werden kann. Einen vergleichbar großen Vorsprung haben die Lehrkräfte auch im Fachwissenstest ( $t(93) = 3.89$ ,  $p < .001$ ,  $d = .73$ ). Vor dem Hintergrund quantitativ wie qualitativ unterschiedlicher Ausbildungsgänge ist zudem davon auszugehen, dass Gymnasiallehrkräfte und -studierende in beiden Testteilen höhere Scores erreichen als ihre nicht gymnasialen Kolleginnen und Kollegen beziehungsweise Kommilitoninnen und Kommilitonen. Die Streudiagramme in Abbildung 4 veranschaulichen entsprechende Performanzunterschiede:



○ / ● Nicht gymnasiales Lehramt; ◇ / ◆: Lehramt Gymnasium; die Gitterlinien kennzeichnen jeweils die Mittelwerte der Gesamtstichprobe.

Abbildung 4: Streudiagramm Fachwissen und fachdidaktisches Wissen (nach Statusgruppen getrennt) mit Skalenmittelwerten der Gesamtstichprobe

Am oberen Ende des Leistungsspektrums finden sich – theoriekonform – ausschließlich Gymnasiallehrkräfte sowie eine Realschulseminarlehrerin (überdurchschnittliche Leistungen in beiden Testteilen, Score „Fachdidaktisches Wissen“ (FDW)



> 44.96 in Kombination mit einem Fachwissensscore (FW) > 20.95); am unteren Ende (FDW-Score < 23.38 in Kombination mit einem FW-Score < 8.91) erscheinen – ebenfalls erwartungsgemäß – nur Studierende nicht gymnasialer Lehrämter:<sup>5</sup>

Ein distales Maß, das sich als Prädiktor für das Abschneiden im FALKO-M-Test eignet, ist die Abiturnote: Zieht man sie als Indikator für allgemeine kognitive Grundfertigkeiten heran (vgl. Baron-Boldt, Schuler & Funke, 1988), so ist die in Tabelle 4 ablesbare Korrelation zwischen Abiturnoten und Testleistungen in beiden Wissensbereichen grundsätzlich erwartungskonform<sup>6</sup> – auch wenn sie bei den Lehrkräften für den FDW-Test überraschend hoch ausfällt:<sup>7</sup>

Tabelle 4: Produkt-Moment-Korrelationen zwischen dem Professionswissen und Indikatoren des Schul-/Studienerfolges beziehungsweise der Berufs-/Studiendauer

N = 129	Abiturnote	Fachsemester	Note des ersten Staatsexamens	Note des zweiten Staatsexamens
Stud. (n = 51)				
FW (8)	-.46**	.64**		
FDW (11)	-.39*	.50**		
Leh. (n = 78)				
		Berufsjahre		
FW (8)	-.20	-.22	-.54**	-.17
FDW (11)	-.57**	.01	-.39**	-.39**

FW = Fachwissen, FDW = fachdidaktisches Wissen. Da einige Testteilnehmerinnen und -teilnehmer fehlende Werte bei einer oder mehreren Variablen hatten, wurden die Korrelationen mithilfe der Full-Information-Maximum-Likelihood-Prozedur geschätzt (Enders, 2010, S. 86–88). Signifikanzniveaus: \*  $p \leq .05$ ; \*\*  $p \leq .01$ .

Die in Tabelle 4 erkennbaren Zusammenhänge zwischen den Gesamtnoten der beiden Staatsexamina und den Testleistungen von Lehrkräften können als Indiz für Kriteriumsvalidität interpretiert werden. Die Korrelationen in mittlerer Größenordnung lassen darauf schließen, dass der Test Wissensbestände erfasst, die sich inhaltlich mit den in den Staatsexamina geprüften Kenntnissen und Fertigkeiten überschneiden. Gemeinsam mit den mittleren bis hohen Korrelationen zwischen den Testergebnissen der Studierenden und den bereits absolvierten Studiensemestern legt dies nahe, dass die Items von FALKO-M tatsächlich professionelles musikpädagogisches Wissen erfassen, das im Laufe des Studiums er-

- 5 Als „überdurchschnittlich“ wird eine Testleistung eingestuft, die über der Summe aus dem Mittelwert und einer Standardabweichung liegt (fachdidaktisches Wissen:  $M = 34.17$ ,  $SD = 10.79$ , Fachwissen:  $M = 14.93$ ,  $SD = 6.02$ ); die Definition einer unterdurchschnittlichen Leistung ergibt sich analog.
- 6 Die Korrelation ist negativ, da im Falle der Abiturnoten kleinere Werte bessere Leistungen indizieren, während es sich bei den Testscores umgekehrt verhält.
- 7 Befunde zur Regression von Fachwissen (FW) und fachdidaktischem Wissen (FDW) auf die Prädiktoren beruflicher Status, studierte Schulform, Abiturnote und Geschlecht: siehe Puffer & Hofmann, 2017, S. 274.

worben wird, und nicht etwa Wissensbestände, über die auch musikinteressierte Laien gewöhnlich verfügen. Die Tatsache, dass die Korrelation zwischen der Note des zweiten Staatsexamens und dem FDW-Score deutlich höher ausfällt als jene mit dem Fachwissensscore, entspricht ebenfalls den Erwartungen: Gegenstand der Zweiten Staatsprüfung sind in Bayern fast ausschließlich unterrichtspraktische und berufsfeldspezifische pädagogische Kompetenzen (Bayerisches Staatsministerium für Unterricht und Kultus, 2004, §§ 18–21).

Dienstältere Lehrkräfte („Berufsjahre“) schneiden im Test nicht besser ab als dienstjüngere – ein Phänomen, das sich in fast allen FALKO-Fächern sowie in der COACTIV-Studie gleichermaßen zeigt und zumindest für den Bereich des fachdidaktischen Wissens theoretischen Postulaten wie Shulmans „wisdom of practice“ zu widersprechen scheint (Lindl & Krauss, 2017, S. 417–419). Da bislang ausschließlich Querschnittsdaten vorliegen, lässt das Fehlen signifikanter Zusammenhänge zwischen Berufsjahren und Professionswissen allerdings nicht den Schluss zu, dass im Verlauf der dritten Phase der Lehrerbildung kein Professionswissen mehr erworben oder aktualisiert würde. Für den Bereich fachdidaktischer Kompetenzen wäre insbesondere zu prüfen, ob in den Jahren nach dem Ende der formalen Ausbildung möglicherweise Zuwächse erfolgen, die sich der Messung mit einem Papier-und-Bleistift-Test entziehen. Für den Bereich des Fachwissens ist zu fragen, inwieweit Kohorteneffekte für den erwartungswidrig fehlenden Leistungsunterschied verantwortlich sein könnten. Für die FALKO-M-Stichprobe ergibt sich beispielsweise eine deutlich ausgeprägte negative Korrelation zwischen der Zahl der absolvierten Berufsjahre und dem Abschneiden in jenen beiden Aufgaben des Fachwissenstests, die Aspekte populärer Musik zum Gegenstand haben (für die aus diesen beiden Items bestehende „Kurzskala“ gilt  $r = -.42^{**}$ ; für die übrigen sechs Fachwissensitems konnte dagegen kein signifikanter Zusammenhang zwischen Berufsjahren und Testleistung ermittelt werden ( $r = .02$ )).

### Dimensionalität des fachspezifischen Professionswissens

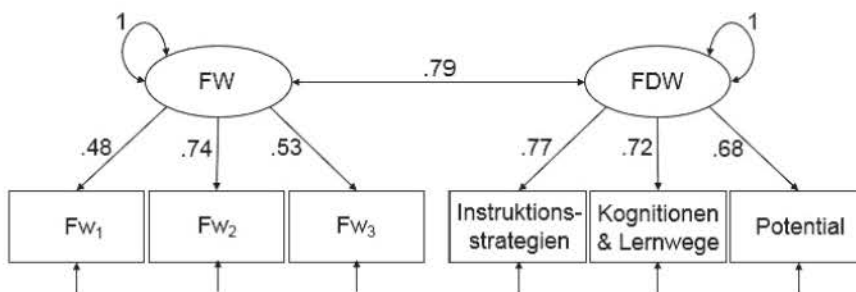
Alle in FALKO-M vorgenommenen Modellierungen sowie die gesamte Testentwicklung und -auswertung basieren auf der Annahme, dass sich professionelles Wissen von Musiklehrkräften mithilfe von Shulmans theoretischen Kategorien „Fachwissen“ und „Fachdidaktisches Wissen“ (Shulman, 1986, 1987) beschreiben lässt. Die Struktur zweier zwar zusammenhängender, aber nicht zusammenfallender Wissensbereiche müsste sich folglich auch in den Daten der Validierungsstudie widerspiegeln. Die im Abschnitt „Psychometrische Gütekriterien“ vorgestellten Kennzahlen sowie die manifeste Korrelation zwischen den Summenscores für fachdidaktisches Wissen und Fachwissen ( $r = .53^{**}$ ) verweisen bereits auf innere Konsistenz und Abgrenzbarkeit der beiden Skalen. Weiteren Aufschluss



über deren Dimensionalität ermöglicht eine konfirmatorische Faktorenanalyse (zur Notwendigkeit der Überprüfung der Dimensionalität: vgl. Bühner, 2011, S. 53). Angesichts der vorliegenden Datenbasis (129 Studienteilnehmerinnen und -teilnehmer, Ad-hoc-Stichprobe) ist die Aussagekraft einer solchen Analyse begrenzt; erste Rückschlüsse bezüglich der Passung zwischen theoretischem Modell und empirischen Daten sind aber möglich.

Im Rahmen der Modellspezifikation wurden die Items des FW- und des FDW-Tests zu je drei Paketen (Parcels) gebündelt, die als manifeste Indikatoren der latenten Konstrukte Fachwissen und Fachdidaktisches Wissen fungieren. Ziel eines solchen Vorgehens ist es, ein akzeptables Verhältnis zwischen Modellkomplexität und Datengrundlage zu erreichen (vgl. z. B. Schermelleh-Engel & Werner, 2009); für die FALKO-Tests wurde ein Verhältnis von 10 Testteilnehmerinnen und teilnehmern pro zu schätzendem Parameter angestrebt. Die Parcels für die FDW-Variable wurden unter inhaltlichen Gesichtspunkten aus den Items der drei FDW-Subskalen gewonnen. Aus den Items des Fachwissenstests wurden ebenfalls drei Pakete gebildet. Da für das Fachwissen in FALKO keine Subfacetten modelliert sind, sind die FW-Parcels nicht auf Grundlage theoretischer Erwägungen zusammengestellt, sondern im Hinblick auf statistische Kriterien: Die drei Pakete sollten möglichst ähnlich, also hinsichtlich Mittelwerten und Standardabweichungen vergleichbar sein.

Abbildung 5 veranschaulicht das Strukturmodell und die Resultate der Schätzungen:



Modellfit:  $\chi^2(8, N = 129) = 14.04$ ;  $p = .08$ ; CFI = .97; RMSEA = .08; SRMR = .05

Abbildung 5: Konfirmatorisches Faktorenmodell des Fachwissens und des fachdidaktischen Wissens (latente Modellierung auf Basis der Gesamtstichprobe  $N = 129$ )

Werte von  $p \geq .01$ , CFI (Comparative Fit Index)  $\geq .95$ , RMSEA (Root-Mean-Square Error of Approximation)  $\leq .08$  und SRMR (Standardized Root Mean Residual)  $\leq .10$  sowie von  $\chi^2 \leq 3df$  (df: Anzahl an Freiheitsgraden) indizieren für Stichprobengrößen  $\leq 250$  akzeptablen Modellfit (Bühner, 2011, S. 418–428; Moosbrugger & Schermelleh-Engel, 2012, S. 337f.).

Die globale Anpassungsgüte des Modells ist positiv zu beurteilen, die Fit-Indizes genügen den in der Literatur vorgeschlagenen Kriterien für die Annahme eines Modells. Die einzelnen Faktorladungen lassen auf einen guten lokalen Fit schließen, die Items beider Testteile laden also tatsächlich auf den postulierten Faktoren. Die beiden latenten Variablen FW (Fachwissen) und FDW (fachdidaktisches Wissen) korrelieren hoch miteinander. Um auszuschließen, dass beide Variablen für dasselbe Konstrukt stehen könnten (FW und FDW würden dann zusammenfallen), wurde ein Vergleichstest durchgeführt zwischen dem Modell in Abbildung 5 und einem alternativen Modell, bei dem die latente Korrelation zwischen FW und FDW auf 1 fixiert wurde. Es ergab sich ein signifikanter Unterschied ( $\Delta\chi^2(1, N = 129) = 7.60; p < .01$ ; zum Vorgehen: vgl. Baltes-Götz, 2015, S. 59–60). Dies kann als Anhaltspunkt dafür gewertet werden, dass die von Shulman postulierte und in FALKO-M für das Fach Musik modellierte Struktur aus zwei überlappenden, aber dennoch substantiell verschiedenen Wissensbereichen „Fachwissen“ und „Fachdidaktisches Wissen“ im FALKO-M-Test auch tatsächlich abgebildet ist.

## Zusammenfassung

Das hier vorgestellte Testinstrument hat angesichts etlicher vorläufiger Setzungen und Behelfslösungen, die im Zuge seiner Entwicklung getroffen beziehungsweise gefunden werden mussten, in vielerlei Hinsicht provisorischen Charakter. Die Ergebnisse der Validierungsstudie legen dennoch nahe, dass die Entwicklung eines ersten reliablen und validen Tests zum Professionswissen von Musiklehrkräften gelungen ist. Damit sind erstmals Messungen dieses Ausschnitts professioneller Kompetenz möglich. Zudem stützen die Arbeitsergebnisse die Annahme, dass Lee Shulmans Modell einer zweifaktoriellen Struktur fachspezifischen Professionswissens (Fachwissen versus fachdidaktisches Wissen) auch auf professionelles Wissen von Musiklehrkräften anwendbar ist. Beide Wissensfacetten ließen sich fachspezifisch konkretisieren, in ein Testkonzept umsetzen und empirisch validieren. Über diesen Beitrag zu musikpädagogischer Theorieentwicklung hinaus hoffen wir, mit unserer Arbeit an FALKO-M auch Anregungen zu weiterer fachdidaktischer Diskussion zu geben.

Zu beachten ist auch, dass der Test nur einen Teilbereich des Professionswissensspektrums von Musiklehrkräften abdeckt. Zum einen wurde er mit Blick auf curriculare Anforderungen an schulischen Musikunterricht in den Jahrgangsstufen 5 und 6 entwickelt; für Messungen von Professionswissensbeständen, die sich auf das Unterrichten in höheren Klassenstufen beziehen, ist das Verfahren nicht geeignet. Zum anderen sei nochmals auf die Vermutung verwiesen, dass es über die im Projekt FALKO untersuchten Professionswissensfacetten hinaus noch weitere geben könnte, die für das Fach Musik Relevanz besitzen – insbesondere im Bereich des fachdidaktischen Wissens (Puffer & Hofmann, 2016,



S. 115). Noch weitgehend ungeklärt sind theoretische Verortung und Messbarkeit der für das Unterrichten von Musik ebenfalls hoch relevanten künstlerisch-musikpraktischen Fähigkeiten und Fertigkeiten. Zudem konnte das Testverfahren bislang nur für bayerische Musiklehrkräfte und -studierende validiert werden; Aussagen über seine Gültigkeit für Studierende und Unterrichtende aus anderen Bundesländern oder für Schulformen, die in Bayern nicht existieren, können derzeit noch nicht getroffen werden.

FALCO-M wurde primär als Messinstrument zur Erkundung von Professionalisierungshypothesen entwickelt. Die prädiktive Validität der Testergebnisse für die Qualität von Musikunterricht sowie für Lernfortschritte von Schülerinnen und Schülern ist noch ungeklärt. Um entsprechende Zusammenhänge zu ergründen, wären deutlich umfangreichere und komplexere Erhebungen erforderlich, für die weitere ökologisch valide Instrumente herangezogen werden müssten (vgl. Weinert & Helmke, 1996; Mulder & Gruber, 2011, S. 435f.; Lindmeier, 2013, S. 12f.). Als Komponente im Rahmen einer solchermaßen umfassenden Studie könnte FALCO-M künftig auch einen Beitrag zur Untersuchung von Wirkungszusammenhängen rund um schulischen Musikunterricht leisten.

## Literatur

- Baltes-Götz, B. (2015). *Analyse von Strukturgleichungsmodellen mit Amos 18*. Universität Trier. <http://www.uni-trier.de/fileadmin/urt/doku/amos/v18/amos18.pdf> [08.04.2017].
- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). Prädiktive Validität von Schulabschlussnoten: Eine Metaanalyse. *Zeitschrift für pädagogische Psychologie*, 2(2), 79–90.
- Baumert, J. & Kunter, M. (2011a). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Baumert, J. & Kunter, M. (2011b). Das mathematikspezifische Wissen von Lehrkräften, kognitive Aktivierung im Unterricht und Lernfortschritte von Schülerinnen und Schülern. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 163–192). Münster: Waxmann.
- Bayerisches Staatsministerium für Unterricht und Kultus (2004). *Ordnung der Zweiten Staatsprüfung für ein Lehramt an öffentlichen Schulen: Lehramtsprüfungsordnung II – LPO II*. [http://www.gesetze-bayern.de/Content/Document/BayLPO\\_II/](http://www.gesetze-bayern.de/Content/Document/BayLPO_II/) [20.12.2016].
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (StMBW) (Hrsg.) (2014). *LehrplanPLUS Grundschule: Lehrplan für die bayerische Grundschule*. <http://www.lehrplanplus.bayern.de/sixcms/media.php/107/LehrplanPLUS%20Grundschule%20StMBW%20-%20Mai%202014.96297.pdf> [20.02.2017].

- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Auflage). Heidelberg: Springer.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Auflage). München: Pearson Studium.
- Dahl, G. (1971). Zur Berechnung des Schwierigkeitsindex bei quantitativ abgestufter Aufgabenbewertung. *Diagnostica*, (17), 139–142.
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York: Guilford Press.
- Gruber, H. & Stöger, H. (2011). Experten-Novizen-Paradigma. In E. Kiel & K. Zierer (Hrsg.), *Unterrichtsgestaltung als Gegenstand der Wissenschaft* (= Basiswissen Unterrichtsgestaltung, Bd. 2) (S. 247–264). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Gruhn, W. (2003). *Geschichte der Musikerziehung: Eine Kultur- und Sozialgeschichte vom Gesangunterricht der Aufklärungspädagogik zu ästhetisch-kultureller Bildung* (2., überarbeitete und erweiterte Auflage). Hofheim: Wolke.
- Jank, W. (Hrsg.) (2013). *MusikDidaktik: Praxishandbuch für die Sekundarstufe I und II* (5., überarbeitete Neuauflage). Berlin: Cornelsen Scriptor.
- Kaiser, H. J. & Nolte, E. (1989). *Musikdidaktik: Sachverhalte – Argumente – Begründungen. Ein Lese- und Arbeitsbuch* (= Schott-Musikpädagogik). Mainz: Schott.
- Kraemer, R.-D. (2007). *Musikpädagogik – eine Einführung in das Studium* (2., verbesserte Auflage) (= Forum Musikpädagogik, Bd. 55). Augsburg: Wißner.
- Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M. & Elsner, J. (2011). Konzeptualisierung und Testkonstruktion zum fachbezogenen Professionswissen von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 135–161). Münster: Waxmann.
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P. & Mulder, R. H. (Hrsg.) (2017). *FALKO: Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik*. Münster: Waxmann.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Lindl, A. & Krauss, S. (2017). Transdisziplinäre Perspektiven auf domänenspezifische Lehrerkompetenzen: Eine Metaanalyse zentraler Resultate des Forschungsprojektes FALKO. In S. Krauss, A. Lindl, A. Schilcher, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 381–438). Münster: Waxmann.



- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (= Fachdidaktische Forschungen, Bd. 4) (S. 45–62). Münster: Waxmann.
- Mayer, J. & Wellnitz, N. (2014). Die Entwicklung von Kompetenzstrukturmodellen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 19–29). Berlin: Springer Spektrum.
- Moosbrugger, H. & Schermelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage) (= Springer-Lehrbuch) (S. 325–343). Berlin/Heidelberg: Springer Berlin Heidelberg.
- Mulder, R. H. & Gruber, H. (2011). Die Lehrperson im Lichte von Professions-, Kompetenz- und Expertiseforschung – die drei Seiten einer Medaille. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven* (S. 427–438). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Niessen, A., Lehmann-Wermser, A., Knigge, J. & Lehmann, A. C. (2008). Entwurf eines Kompetenzmodells ‚Musik wahrnehmen und kontextualisieren‘. *Zeitschrift für Kritische Musikpädagogik*, 3–33. <http://www.zfkm.org/sonder08-niessenetal.pdf> [20.12.2016].
- Puffer, G. & Hofmann, B. (2016). FALKO-M: Zur Konzeptualisierung des Professionswissens von Musiklehrkräften. In J. Knigge & A. Niessen (Hrsg.), *Musikpädagogik und Erziehungswissenschaft* (= Musikpädagogische Forschung, Bd. 37)(S. 107–120). Münster: Waxmann.
- Puffer, G. & Hofmann, B. (2017). FALKO-M: Entwicklung und Validierung eines Testinstruments zum domänenspezifischen Professionswissen von Musiklehrkräften. In S. Krauss, A. Schilcher, A. Lindl, M. Fricke, A. Göhring, B. Hofmann, P. Kirchhoff & R. H. Mulder (Hrsg.), *FALKO: Fachspezifische Lehrerkompetenzen: Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Evangelische Religionslehre, Musik und Pädagogik* (S. 245–289). Münster: Waxmann.
- Schermelleh-Engel, K. & Werner, C. (2009). *Item Parceling: Bildung von Testteilen oder Item-Päckchen. Analyse von Strukturgleichungsmodellen mit LISREL*. [http://www.psychologie.uzh.ch/dam/jcr:ffffff-b371-2797-0000-00000ed9f491/item\\_parceling.pdf](http://www.psychologie.uzh.ch/dam/jcr:ffffff-b371-2797-0000-00000ed9f491/item_parceling.pdf) [08.04.2017].
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Venus, D. (1969). *Unterweisung im Musikhören* (= Musikpädagogische Bibliothek, Bd. 30). Wilhelmshaven: Noetzel.
- Voss, T., Kunina-Habenicht, O., Hoehne, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187–223.

Weinert, F. E. & Helmke, A. (1996). Der gute Lehrer: Person, Funktion oder Fiktion? In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen* (= Zeitschrift für Pädagogik: Beiheft) (S. 223–233). Weinheim/Basel: Beltz.

Dr. Gabriele Puffer  
Universität Augsburg  
Lehrstuhl für Musikpädagogik  
Universitätsstr. 26  
86159 Augsburg  
gabriele.puffer@phil.uni-augsburg.de